Documentation projet InvestQ

# Liste de ressources utilisées

* [Exemple de projet d’apprentissage par renforcement](https://medium.datadriveninvestor.com/how-to-use-reinforcement-learning-for-profitable-investing-ebe8662f13cc)
* [Documentation de l’API Alpaca](https://docs.alpaca.markets/reference/stockbars-1)
* [Article 2018](https://arxiv.org/pdf/1811.07522)

# Notions et vocabulaires

## Titre

Un titre(stock), également connue sous le nom « equity », est un titre financier qui représente la propriété d'une fraction de la société émettrice. Les unités d'actions sont appelées parts (shares), qui donnent à leur propriétaire droit à une proportion des actifs (assets) et des bénéfices de la société correspondant à la quantité d'actions qu'il possède.

Les actions sont principalement achetées et vendues sur les bourses et constituent la base des portefeuilles de nombreux investisseurs individuels. Les transactions boursières doivent respecter les réglementations gouvernementales destinées à protéger les investisseurs contre les pratiques frauduleuses.

## Positions

Une position est la quantité d'un titre, d'un actif ou d'une propriété possédée (ou vendue à découvert) par un individu ou une autre entité. Un trader ou un investisseur prend une position lorsqu'il effectue un achat via un ordre d'achat, indiquant une intention haussière ; ou lorsqu'il vend des titres à découvert avec une intention baissière.

L'ouverture d'une nouvelle position est finalement suivie à un moment donné dans le futur par la sortie ou la clôture de la position (c’est-à-dire la vente de tous les parts).

### Position longue

Une position longue signifie que l'investisseur a acheté un actif financier (comme des actions, des obligations ou des marchandises) et s'attend à ce que le prix de cet actif augmente.

L'objectif est de vendre l'actif à un prix plus élevé dans le futur pour réaliser un profit.

### Position courte

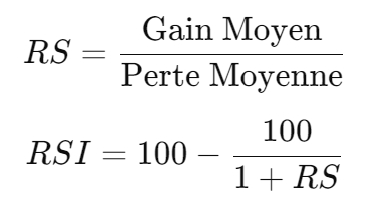
Une position courte signifie que l'investisseur a emprunté un actif financier et l'a vendu avec l'intention de le racheter plus tard à un prix inférieur.

L'objectif est de profiter de la baisse du prix de l'actif en le rachetant à un prix inférieur à celui de la vente initiale.

# Indicateurs techniques d’analyse

## L’indicateur RSI

L'indice de force relative (RSI) est un indicateur technique populaire utilisé dans l'analyse des marchés financiers. Il est conçu pour mesurer la vitesse et le changement des mouvements de prix. Le RSI oscille entre 0 et 100 et est généralement utilisé pour identifier les conditions de surachat ou de survente dans un marché.



### L’indicateur MACD

Le MACD (Moving Average Convergence Divergence) est un indicateur technique utilisé dans l'analyse technique des marchés financiers. Il est basé sur les différences entre deux moyennes mobiles exponentielles (EMA) et est souvent utilisé pour identifier les tendances et les points d'inversion potentiels.

### L’indicateur ADX

L'indice directionnel moyen (ADX) est un indicateur d'analyse technique utilisé par certains traders pour déterminer la force d'une tendance.

La tendance peut être soit à la hausse soit à la baisse, et cela est montré par deux indicateurs d'accompagnement, l'indicateur directionnel négatif (-DI) et l'indicateur directionnel positif (+DI). Par conséquent, l'ADX comprend généralement trois lignes distinctes.

2 indicateurs de performances :

* Sharpe ratio
* Cumulative returns

# Récupération des données historiques

L’historique d’Alpaca remonte jusqu’à 2016 mais avec des données manquantes. On utilisera les données depuis 2017 afin qu’elles soient plus complète.

4 types de données historiques :

* Historical bars : correspond aux candlesticks qu’on va détailler ci-dessous
* Historical auctions : …
* Historical quotes : …
* Historical trades : …

**SIP** is short for [Securities Information Processor](https://en.wikipedia.org/wiki/Securities_information_processor). All US exchanges are mandated by the regulators to report their activities (trades and quotes) to the consolidated tape. This is what we call SIP data.

**IEX** ([Investors Exchange](https://en.wikipedia.org/wiki/IEX)) is a single stock exchange.

On utilisera le SIP pour les données historiques et IEX pour récupérer les données en temps réel étant donné qu’on n’a pas accès aux données SIP en temps réel avec la version gratuite de l’API Alpaca.

On récupère une liste de stocks avec 2 times frame différents : L’état du stock sur une durée de 1 jour et sur 1 heure.

API limitée à 200 requêtes par minute => il faut donc ajouter un sleep pour éviter un blocage de l’API surtout pour un time frame d’une heure (beaucoup plus de données et donc d’appel à l’API)

On récupère des données sous la forme de « candlestick ». Cette forme permet de représenter l’état du marché durant l’intervalle de temps défini, en indiquant :

* Le prix d’ouverture -> la valeur de la première transaction
* Le prix de fermeture -> la valeur de la dernière transaction
* Le prix le plus élevé durant cet intervalle de temps
* Le prix le plus bas durant cet intervalle de temps

Ajouté à ça :

* Le nombre d’échange effectué
* L’horodatage indiquant le début de période spécifiée
* Le volume d’actions échangées durant la période spécifiée
* Le Volume-Weighted Average Price - VWAP : Le prix moyen pondéré par le volume pour la période spécifiée. Il est calculé en multipliant chaque prix de transaction par le volume de cette transaction, puis en divisant la somme des produits par le volume total des transactions

# Etat de l’art

Cet état de l’art est basé sur l'article de Ying Yu publié en 2024 [[1](https://www.atlantis-press.com/proceedings/icbis-24/125999560)].

L'utilisation du machine learning dans les marchés financiers n'est pas nouvelle. Cependant, les grandes avancées de l'apprentissage par renforcement depuis 2016 ont relancé cet axe de recherche. Avant 2016, les articles sur ce sujet étaient rares (quelques articles avant les années 2000), avec une période de vide notable avant un regain d'intérêt récent.

Ces algorithmes sont utilisés dans plusieurs applications :

1. **Optimisation des stratégies de trading** :
   * Ils déterminent les meilleures actions à prendre dans diverses conditions de marché.
   * Ces modèles peuvent ajuster les portefeuilles en temps réel, maximisant les rendements tout en minimisant les risques.
2. **Gestion des risques** :
   * Les agents d'apprentissage par renforcement (RL) peuvent être programmés pour adopter des comportements conservateurs ou agressifs en fonction des conditions du marché, aidant ainsi à réduire les pertes potentielles.
   * Par exemple, ils peuvent définir des points de stop-loss et de take-profit pour limiter les risques liés aux fluctuations du marché. Les stop-loss et take-profit sont des valeurs seuils définies pour vendre automatiquement une action lorsque ces seuils sont atteints.
3. **Arbitrage** :
   * Les modèles RL identifient et exploitent les opportunités d'arbitrage entre différents instruments financiers, tels que les contrats à terme avec des dates d'expiration différentes.
   * Cela permet de réaliser des profits en exploitant les inefficacités du marché sans prendre de risques directionnels significatifs.
4. **Capacités prédictives** :
   * Les modèles de machine learning analysent les données historiques pour prévoir les mouvements futurs des prix des actions.
   * Ces prévisions se basent sur l'analyse d'indicateurs techniques, de données fondamentales et d'autres informations pertinentes pour anticiper les tendances du marché.

On distingue deux grandes catégories de stratégies de trading :

1. **Stratégies à haute fréquence** :
   * Elles exploitent de petites inefficiences de marché sur de courtes périodes, souvent de l'ordre de millisecondes à quelques secondes.
2. **Stratégies classiques** :
   * Elles visent à optimiser les gains en étudiant les tendances des marchés et en répartissant le portefeuille de manière efficace.

Ces stratégies sont étudiées soit par le biais de prévisions (forecasting) pour prendre des décisions en fonction des prédictions, soit par l'établissement de stratégies de trading à partir de l'apprentissage par renforcement. La tendance actuelle favorise davantage l'établissement de stratégies de trading.

Pour améliorer la précision des modèles, il peut être très utile de coupler l'étude des prix du marché avec l'analyse de sentiments provenant des réseaux sociaux et des nouvelles financières, en particulier pour le trading non basé sur la haute fréquence.

# Expérimentation

Lorsqu'on observe l'évolution du prix d'une action au fil du temps, des tendances distinctes se dessinent. Il serait tentant de croire qu'il suffit d'acheter lorsque le prix augmente et de vendre lorsqu'il diminue sur des périodes suffisamment longues. Pour démontrer la complexité de ce problème, nous allons simuler, à l'aide de données historiques, deux stratégies simples. La première se base sur la différence de prix entre deux instants différents, tandis que la seconde utilise le calcul du momentum pour lisser les variations et mieux faire ressortir les tendances. Ces stratégies ne sont pas destinées à être utilisées en pratique ; elles servent uniquement à illustrer que le problème n'est pas aussi simple qu'il n'y paraît.

Par la suite, nous appliquerons une approche basée sur le machine learning en utilisant des techniques de prévision sur des séries temporelles. Nous utiliserons un modèle basé sur des couches LSTM. Enfin, nous testerons des algorithmes d’apprentissage par renforcement, tels que PPO, connus pour leur efficacité dans des environnements stochastiques.

Les résultats présentés ci-dessous se basent sur les données de l'action Apple entre le 1er janvier 2023 et le mois de mai 2024 (soit une période d'un an et demi). Les valeurs sont enregistrées avec une fréquence horaire (time frame d'une heure). Pour simplifier le problème, nous avons restreint les transactions à l'achat d'une seule action à la fois, avec un portefeuille initial de 10 000 €. Pour l'entraînement des modèles de machine learning, nous utiliserons les données de janvier 2017 à janvier 2022, celles de l’année 2022 pour la validation, et le jeu de test couvrira la période du 1er janvier 2023 à mai 2024. Il convient de noter qu'Apple a divisé le prix de son action par 4 le 31 août 2020 ; nous multiplierons donc les prix par 4 pour éviter de perturber l'apprentissage de nos modèles.

## Approche simple

### Différence de prix

Nous allons commencer par calculer la différence de prix entre deux instants temporels que nous ferons varier. L'idée sous-jacente est que, dans une tendance haussière, le prix actuel doit être supérieur au prix précédent, ce qui suggère un achat. À l'inverse, dans une tendance baissière, le prix actuel devrait être inférieur au prix précédent, ce qui indiquerait une vente. Nous ferons également varier l'intervalle de temps pour ce calcul de différence de prix, car une période plus longue pourrait être plus représentative en raison de la forte volatilité des prix.

Pour évaluer les résultats, nous utiliserons l'évolution de l'action comme ligne de base (baseline).

Une image contenant texte, capture d’écran, Police, Tracé

Description générée automatiquement

Les résultats sont présentés pour des intervalles de différence de prix calculés sur 1, 10, 20, 40, et 80 valeurs. Ces intervalles ne correspondent pas exactement à leur équivalent en heures, car le marché n'est pas ouvert 24 heures sur 24.

Comme le montre le graphique, seule la courbe verte (correspondant à un intervalle de 40) dépasse la performance de l’action à la fin de la période historique, mais le gain n'est pas significativement supérieur. Par conséquent, nous ne pouvons pas conclure que cette stratégie de trading surperforme par rapport à une simple détention à long terme de l'action.

### Lissage du prix par le Momentum

Pour mettre en évidence les tendances, nous allons lisser les fluctuations de prix en utilisant le concept de momentum.



Avec pt le prix au temps t, beta un paramètre entre 0 et 1, et mt le momentum calculé.

On va utiliser la même stratégie qu’avant pour calculer la différence de prix entre le prix actuel et la valeur précédente mais basée sur le momentum du prix. On choisit les valeurs béta suivantes [0.9, 0.95, 0.99, 0.995, 0.999]. Plus la valeur est grande est plus le prix sera lissé mais le momentum sera davantage en retard sur le prix réel.

En prenant **p**t​​ comme le prix à l'instant **t**, β comme un paramètre de lissage compris entre 0 et 1, et **m**t ​ comme le momentum calculé, nous appliquons une stratégie similaire à la précédente pour calculer la différence de prix, mais cette fois basée sur le momentum. Nous testons différentes valeurs de β : [0,9, 0,95, 0,99, 0,995, 0,999]. Plus la valeur de β est élevée, plus le prix sera lissé, mais le momentum accusera un plus grand retard par rapport au prix réel.

Une image contenant texte, capture d’écran, Police, ligne

Description générée automatiquement

Les meilleurs résultats sont obtenus avec un β de 0,99 et 0,995. Cette approche semble plus efficace que la précédente, mais les gains restent trop faibles pour conclure à une réelle efficacité. En effet, pour d'autres périodes ou actions, les paramètres optimaux pourraient varier, et rien ne garantit que l'on obtiendra de meilleurs résultats.

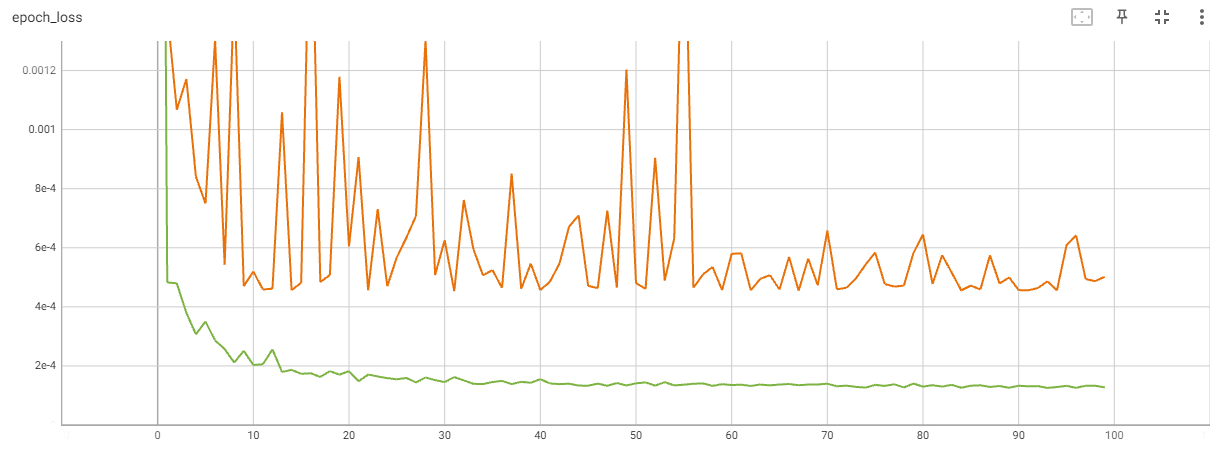
Ces deux expérimentations illustrent la complexité du problème. Effectuer du trading uniquement en se basant sur la tendance donne généralement des résultats inférieurs à une stratégie d'achat à long terme sans trading actif. Nous avons pu optimiser les paramètres pour obtenir des résultats légèrement supérieurs à l'évolution de l'action, car nous connaissions le comportement du prix sur toute la période de trading. Dans une situation réelle où l'évolution future des prix est inconnue, il est impossible de déterminer les meilleurs paramètres à l'avance.

Les traders utilisent de nombreux indicateurs plus sophistiqués, s'appuient sur leur expérience personnelle, et considèrent les événements externes qui pourraient influencer le cours des actions au-delà de la simple analyse des prix.

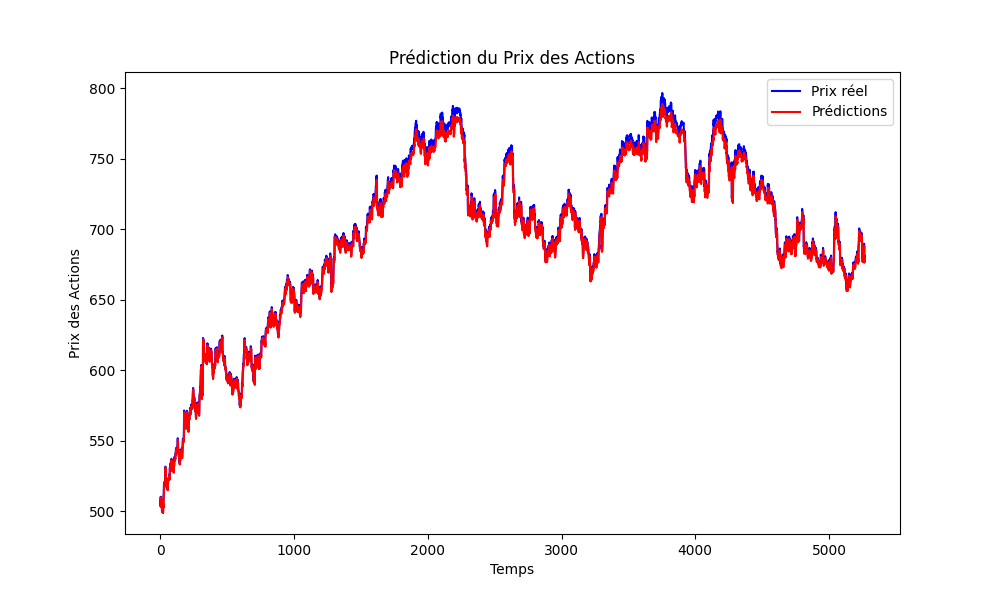
## Approche machine learning

### Approche des séries temporelles avec des réseaux de neurones récurrents LSTM

Les réseaux de neurones récurrents (RNN) sont particulièrement adaptés à la modélisation des séries temporelles. Nous avons testé un modèle simple constitué de deux couches LSTM de 128 neurones, utilisant uniquement les prix de clôture en entrée. La fenêtre d'observation est de 30 valeurs, et nous avons appliqué une normalisation standard aux données en raison de la magnitude élevée des prix.

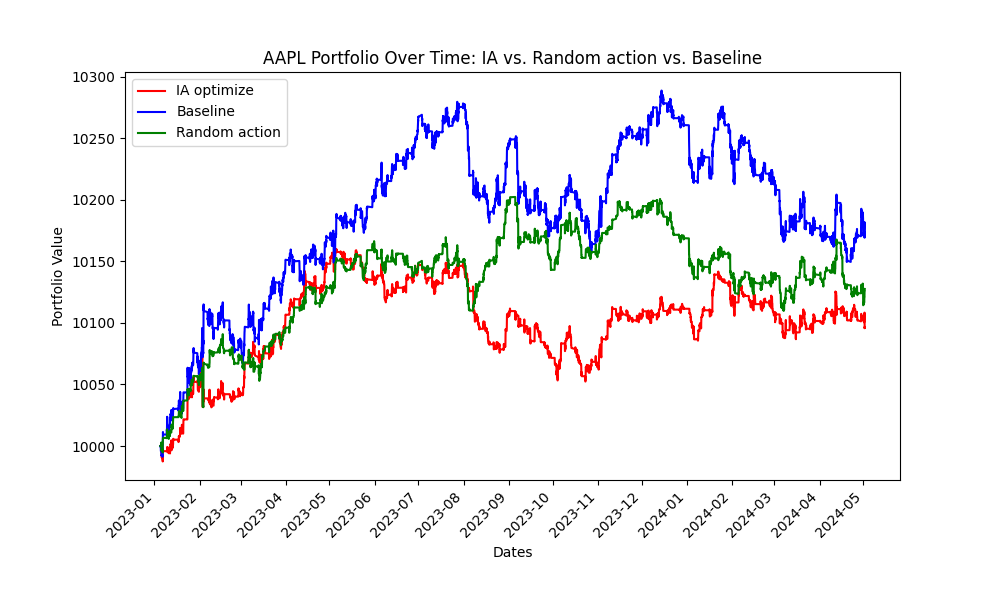


Le modèle a bien convergé, bien qu'un léger surapprentissage soit observé (la courbe de validation en orange dépasse légèrement celle de l'entraînement en vert).



Le graphique ci-dessus compare le prix réel au prix prédit. Nous prédisons le prix à l'instant suivant en utilisant les 30 derniers prix réels. Les résultats semblent bien s'ajuster, mais cela ne prouve pas que le modèle peut effectivement prédire le prix suivant. En effet, le modèle peut simplement prédire une valeur proche de la dernière valeur connue pour obtenir un bon ajustement apparent.

Cela devient évident lorsque l'on calcule la proportion de bonnes prédictions de la direction du prix (hausse ou baisse). Si le modèle prédit une augmentation du prix par rapport au dernier point et que le prix augmente effectivement, on considère cela comme une bonne prédiction, et de même pour les baisses. Pour ce modèle, nous obtenons un taux de 49,3 % de bonnes prédictions, ce qui est essentiellement aléatoire, étant donné qu'il y a une chance sur deux de se tromper (soit 50 %).

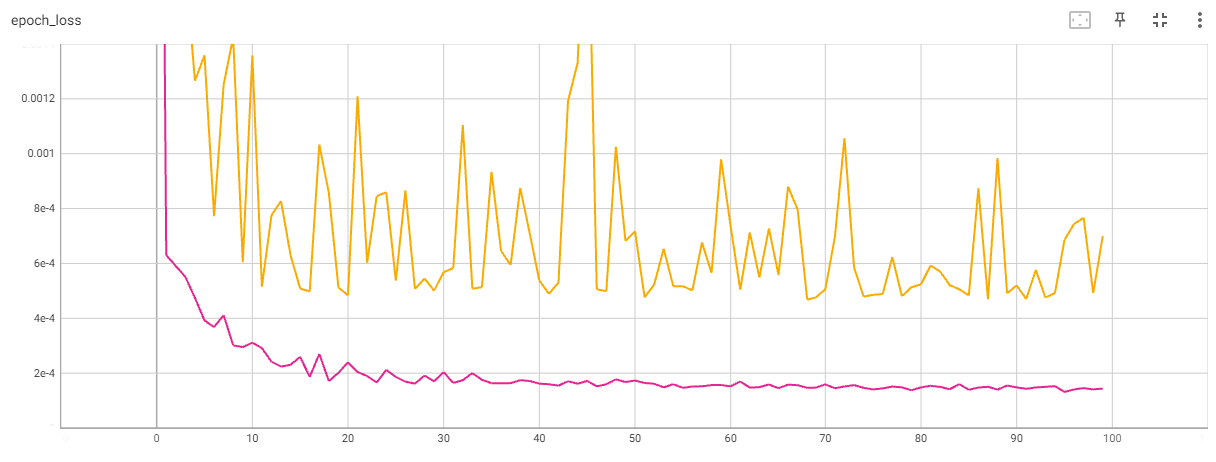


Le graphique ci-dessus montre l'évolution du portefeuille de l'IA par rapport à l'évolution de l'action et des actions aléatoires. Si l'IA prédit une augmentation du prix, elle achète ; sinon, elle vend. Comme mentionné précédemment, bien que le modèle puisse prédire une valeur proche de la valeur réelle, il ne peut pas correctement prévoir si elle sera supérieure ou inférieure à la valeur précédente. Ainsi, les performances de l'IA sont similaires à celles d'une stratégie aléatoire.

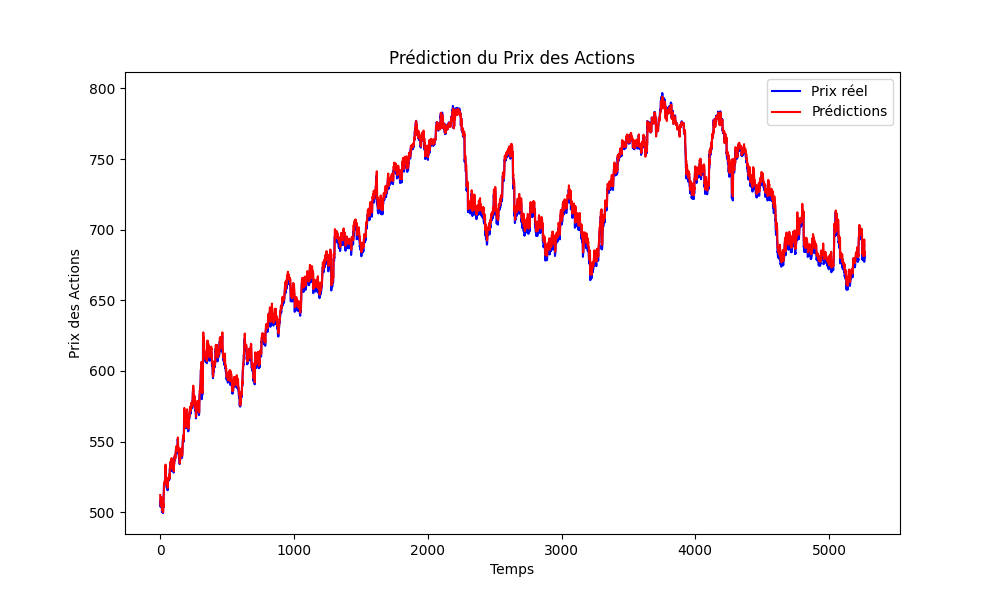
**Ajout de nouvelles fonctionnalités : nombre de transactions et volume échangé**

Pour enrichir notre modèle, nous avons ajouté, en plus des prix de clôture, le nombre de transactions et le volume total des actions échangées. L'objectif est de tester si ces nouvelles caractéristiques (features) peuvent aider notre modèle à mieux comprendre l'évolution des prix.

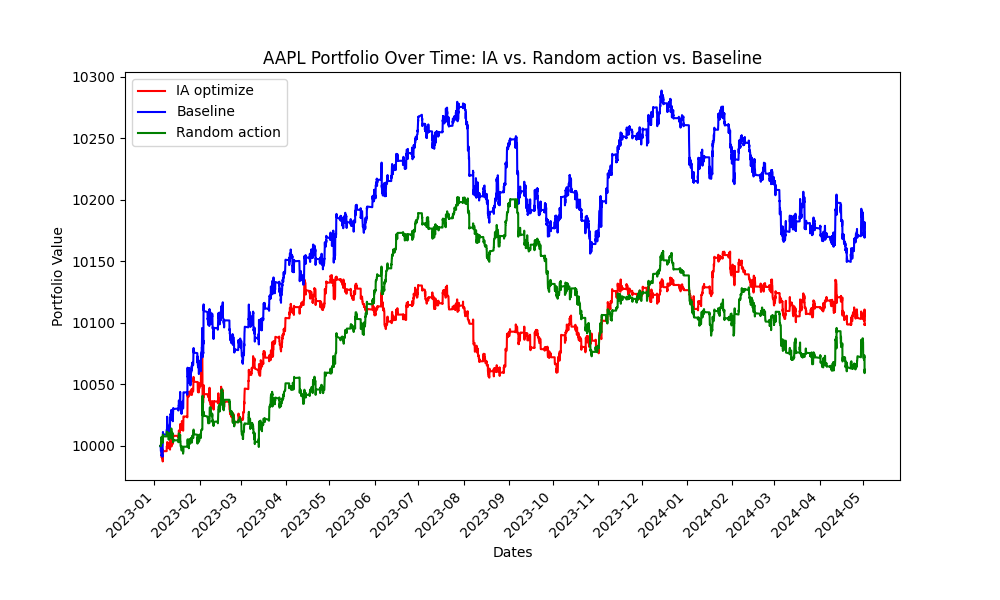
Les mêmes types de graphiques sont représentés ci-dessous, montrant la comparaison entre le prix réel et le prix prédit, ainsi que l'évolution du portefeuille basé sur ces prédictions.



Le modèle a bien convergé, bien qu'un léger surapprentissage soit observé (la courbe de validation en orange dépasse légèrement celle de l'entraînement en rose).



Les résultats montrent un taux de 48,6 % de bonnes prédictions de tendance, ce qui est comparable au modèle précédent. Comme on peut le constater, l'ajout de ces deux nouvelles fonctionnalités n'a pas amélioré la capacité du modèle à prédire la direction des prix.

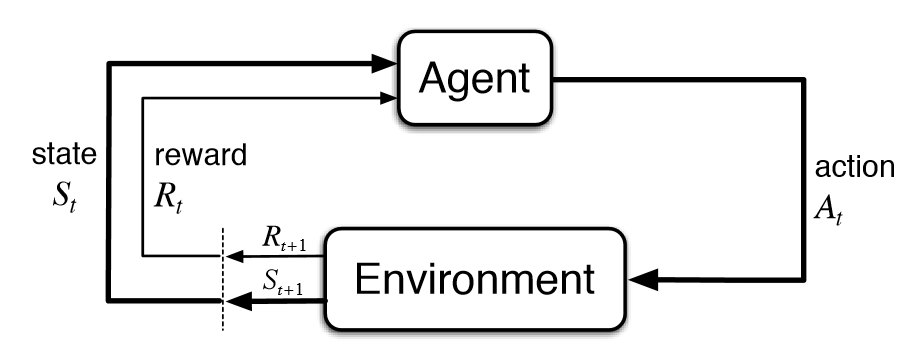


Ces observations confirment que ces deux caractéristiques supplémentaires n'ont pas permis au modèle de mieux saisir l'évolution des prix. En d'autres termes, l'inclusion du nombre de transactions et du volume échangé n'a pas apporté d'amélioration significative à la performance du modèle en termes de prédiction de tendance.

### Apprentissage par renforcement

Avec les réseaux de neurones récurrents, nous avons essayé de prédire l'évolution des prix pour prendre des décisions d'achat ou de vente d'actifs. En revanche, avec les modèles d'apprentissage par renforcement (RL), nous allons demander à notre modèle de déterminer directement l'action à effectuer : acheter, vendre, ou ne rien faire.

Le schéma ci-dessous illustre le principe de l'apprentissage par renforcement :



Dans ce cadre, le modèle ou agent exécute une action sur l'environnement en se basant sur son observation de celui-ci. Au temps suivant, l'environnement fournit une nouvelle observation ainsi que la récompense associée à l'action prise. En utilisant cette nouvelle observation, l'agent effectue une nouvelle action, et ce processus se répète. La récompense sert de feedback à l'agent concernant l'action réalisée, ce qui lui permet d'apprendre. L'objectif de l'agent est de maximiser la somme cumulée des récompenses au fil du temps.

Pour notre problématique, nous visons à maximiser la valeur de notre portefeuille au fil du temps. Ainsi, la récompense sera définie par la variation de la valeur du portefeuille. L'environnement sera représenté par l'évolution des prix des actions. Quant à l'observation, elle peut être construite de plusieurs manières :

* Utiliser uniquement le dernier prix disponible.
* Ajouter des indicateurs calculés sur une période passée pour fournir au modèle une idée des tendances de prix.
* Intégrer l'historique des prix, comme dans le modèle LSTM précédent.
* Inclure des informations supplémentaires telles que le nombre d'actions détenues, la valeur actuelle du portefeuille, etc.

En combinant ces éléments, l'agent est mieux équipé pour prendre des décisions qui maximisent la valeur de notre portefeuille sur le long terme.

**Simplification et Entraînement du Modèle**

Pour simplifier le problème, nous limitons le modèle à la possession d'une seule unité d'action à la fois. Le réseau de neurones utilisé est constitué de deux couches denses de 64 unités chacune.

**Première tentative**

En entrée, nous fournissons l'historique des 30 valeurs de prix précédentes ainsi que la position actuelle, c'est-à-dire si l'agent détient ou non une action. Le modèle est entraîné sur 1 millions de pas (steps), ce qui correspond à 1 millions d'actions effectuées sur le jeu d'entraînement.

Une image contenant texte, ligne, Tracé, Police

Description générée automatiquement

La figure ci-dessous montre l'évolution de la récompense moyenne (reward) au cours des rollouts de 2048 pas. Un rollout correspond à la quantité de données collectées avant d'effectuer une itération d'apprentissage. Étant donné l'instabilité des résultats, nous lissons la courbe pour faire apparaître la tendance générale.

La récompense moyenne augmente mais le modèle ne semble pas avoir fini de converger. Pour visualiser les performances de notre modèle, nous analysons l'évolution du portefeuille sur le jeu de test en appliquant notre modèle entraîné, comme illustré dans le graphique suivant.

Une image contenant texte, capture d’écran, ligne, Tracé

Description générée automatiquementLors de notre première tentative, le modèle a atteint des performances similaires à l’évolution du cours de l’action, ce qui est déjà une net amélioration par rapport au forcasting de timeseries avec des réseaux récurrents. La courbe rouge dans le graphique représente les performances de notre modèle, tandis que la courbe bleue montre une simulation d'une position longue sur toute la durée du test.

**Deuxième tentative**

Pour la deuxième tentative, nous avons adopté une approche plus sophistiquée en fournissant au modèle non seulement la dernière valeur de l'action et la position actuelle, mais également les indicateurs techniques suivants :

* **RSI (Relative Strength Index)**
* **ADX (Average Directional Index)**
* **MACD (Moving Average Convergence Divergence)**

Ces indicateurs sont calculés sur des périodes d'une heure étant donné que nos données historiques sont basées sur un timeframe d’une heure. Par exemple, le RSI est calculé sur une période de 14 heures.

Une image contenant texte, ligne, Tracé, Police

Description générée automatiquement

Avec cette approche, la récompense moyenne a effectivement augmenté au cours d'un total de 1 million de pas (steps) mais le modèle n’est pas encore arrivé à convergence.

Une image contenant texte, ligne, Tracé, Police

Description générée automatiquement

Pour valider ces résultats, nous avons effectué une simulation sur le jeu de test. Notre modèle a obtenu de moins bon résultats sur l’évolution du portfolio que le modèle précédent et on est bien moins performant que notre baseline.

**Troisième tentative**

Afin de donner une vision plus globale à notre modèle, nous avons ajouté les mêmes trois indicateurs avec un timeframe d’un jour. Pour chaque instant t, nous disposons du prix, de la position, et des trois indicateurs avec un timeframe d’une heure et d’un jour.

Une image contenant ligne, Tracé, texte, diagramme

Description générée automatiquement

La récompense moyenne a atteint des valeurs plus élevées avec un nombre de steps équivalent. Le modèle n’a pas totalement convergé, et la récompense pourrait potentiellement être encore plus élevée.

Une image contenant texte, Tracé, ligne, diagramme

Description générée automatiquement

L’évolution du portfolio de test montre d'excellentes performances, surpassant de loin notre baseline, qui correspond à l'évolution du prix de l'action sur la même période. Nous avons relancé l'entraînement une deuxième fois pour confirmer que ces résultats ne sont pas dus au hasard.

Il est important de noter que, malgré ces performances prometteuses, nous devrons tester notre modèle sur des données de marché en temps réel avant de valider définitivement ses capacités.

Dans les prochaines étapes, plutôt que de continuer à expérimenter d'autres indicateurs ou à affiner le modèle pour des gains supplémentaires, nous nous concentrerons sur l'investissement simultané dans plusieurs types d'actions.

### Gestion de Plusieurs Actions

Afin de diversifier les investissements financiers, il est crucial de ne pas tout miser sur une seule action, mais de répartir son capital sur un ensemble d'actifs choisis judicieusement. Pour l'instant, nous nous basons sur les 30 actions différentes de l'indice DJIA. À l'avenir, nous pourrions envisager un modèle capable de sélectionner les actions les plus performantes du moment. Nous avons choisi cette approche pour utiliser la même base d'actions que dans cet article [[1](https://arxiv.org/pdf/1811.07522)].

Liste des symboles : ["AMZN", "AXP", "AMGN", "AAPL", "BA", "CAT", "CSCO", "CVX", "GS", "HD", "HON", "IBM", "INTC", "JNJ", "KO", "JPM", "MCD", "MMM", "MRK", "MSFT", "NKE", "PG", "TRV", "UNH", "CRM", "VZ", "V", "WMT", "DIS", "DOW"]

**Traitement des Données**

Pour traiter les données, nous avons ajusté les prix en supprimant les effets des divisions d'actions pour AMZN, AAPL et WMT. Par exemple, AAPL a été divisé par trois à partir du 31 août 2020, nous avons donc multiplié par trois sa valeur après cette date.

Nous avons calculé les mêmes indicateurs techniques (RSI, ADX, et MACD) pour chacune des actions.

Les dates des données ne concordent pas toutes entre elles en raison des horaires de négociation variables et des heures de négociation prolongées pour différentes actions. Pour harmoniser nos données, nous avons effectué une jointure externe (outer join) lors de la concaténation des 30 actions afin de conserver toutes les dates et de ne pas perdre de données. Les valeurs manquantes ont été remplies par la dernière valeur disponible, car les prix des actions n'évoluent pas lorsque le marché est fermé.

**Features**

Nous voulons un modèle unique capable de gérer les 30 actions simultanément. Pour ce faire, nous avons concaténé toutes les features pour chaque action. En plus de ces features, nous avons ajouté le nombre de parts détenues, la montant équivalent en $, et le montant en cash disponible pour effectuer des achats. La sortie du modèle correspondra aux actions de vendre, acheter ou ne rien faire pour chaque action.

Nous avons initialisé notre environnement avec 100 000 $ de cash. Chaque achat est effectué pour un montant fixe de 1 000 $, et la vente se fait pour l'intégralité des parts détenues de l'action. Si le montant en cash disponible est insuffisant pour un achat, aucune action n'est prise. Pour optimiser le cash, nous effectuons toutes les ventes en premier afin de récupérer le cash nécessaire avant d'effectuer des achats.

**Calcul de la Baseline**

Pour calculer la baseline, nous déterminons le prix moyen de toutes les actions et achetons un nombre de parts équivalent à 100 000 $ au temps initial. L’évolution de la baseline sera simplement le produit entre le nombre de parts et l’évolution du prix moyen.

**Entraînement et Résultats du Modèle**

Une image contenant ligne, diagramme, Tracé, Parallèle

Description générée automatiquement

Nous avons entraîné notre modèle sur 400 000 steps. La récompense (reward) augmente, ce qui montre que notre modèle parvient à apprendre. Cependant, le modèle n’a pas encore convergé et nécessiterait un nombre de steps plus élevé.

Une image contenant texte, capture d’écran, Tracé

Description générée automatiquement

La courbe en bleu est notre baseline avec l’investissement des 100 000$. La courbe rouge l’optimisation du portfolio par l’IA et la courbe noire est le montant en cash non investie par l’IA.

Notre IA performe moins bien que notre baseline mais on remarque qu’elle est très stable. Cela est dû d’une part au manque d’entrainement de notre modèle mais surtout du montant en cash non investi qui représente environ 70% du montant total.

Il va falloir palier à ce problème en forçant l’investissement de tout le cash restant. Pour se faire, pour chaque d’action d’achat on va en effectuer un nouveau jusqu’à qu’il n’y est plus suffisant de cash.

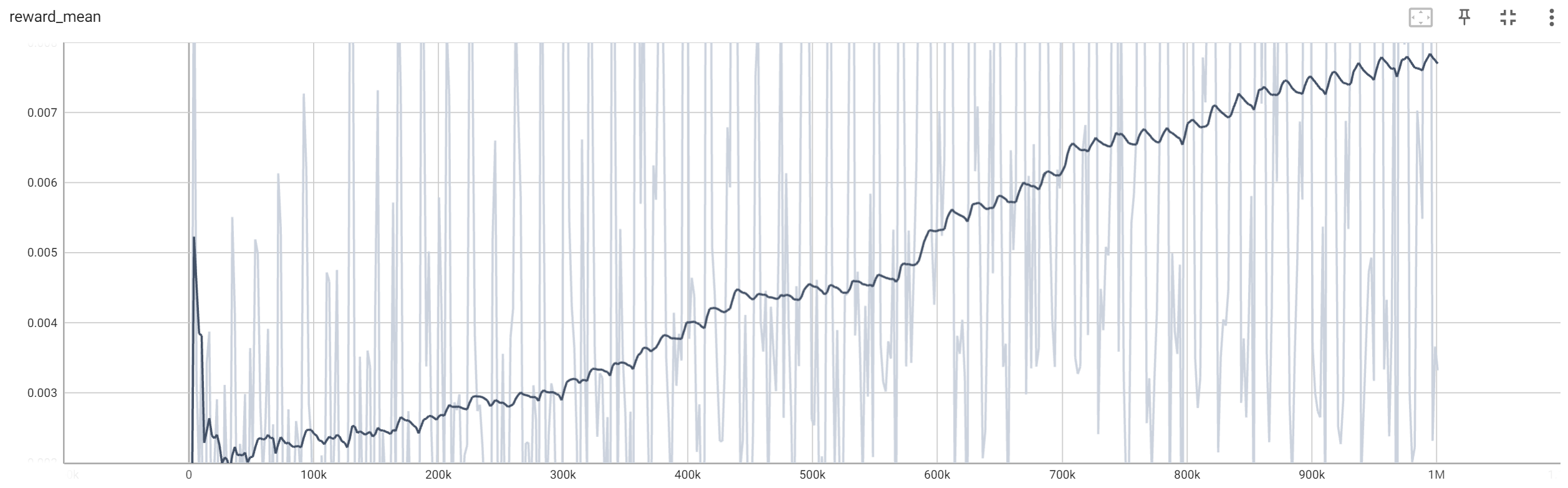
Dans le graphique ci-dessus :

* La courbe bleue représente notre baseline avec un investissement initial de 100 000 $.
* La courbe rouge illustre l’optimisation du portefeuille par l’IA.
* La courbe noire montre le montant en cash non investi par l’IA.

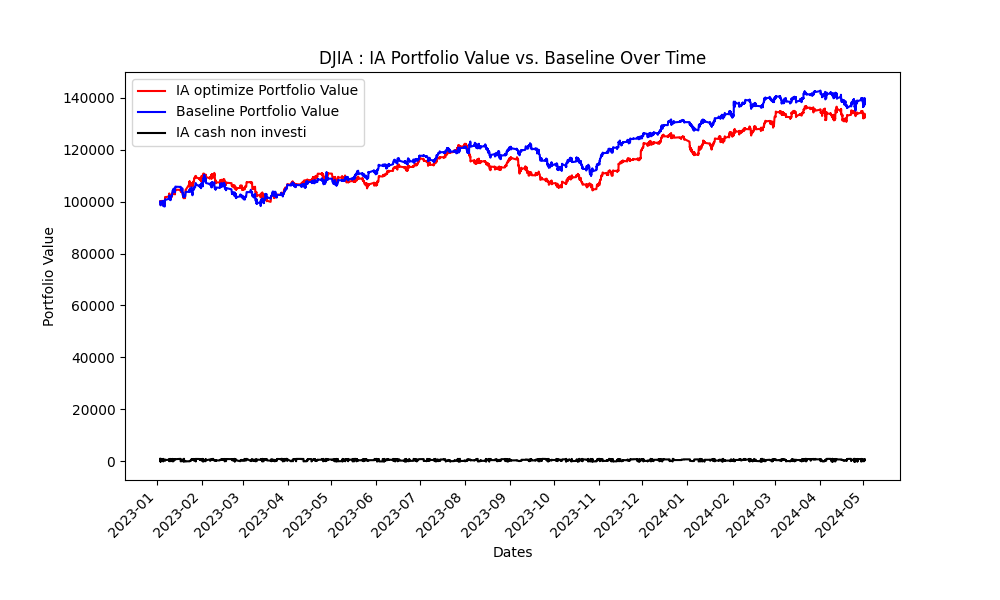
Notre IA performe moins bien que la baseline, mais elle est très stable. Cette sous performance est en partie due au manque d'entraînement du modèle, mais surtout au montant en cash non investi, qui représente environ 70 % du montant total.

**Optimisation de l'Investissement**

Pour résoudre ce problème, nous devons investir la totalité du cash restant. Pour ce faire, après chaque action d'achat, nous en effectuerons une nouvelle jusqu'à épuisement du cash disponible. Nous avons également augmenté la taille du réseau de neurones, passant de 2 couches de 64 neurones chacune à 3 couches de 256, 256 et 64 neurones, pour gérer un total de 511 features.



Après 1 million de pas (environ 2 heures d'entraînement), le modèle n'a pas encore convergé. Nous n'allons pas prolonger l'entraînement car nous en sommes encore à une phase expérimentale.



La courbe noire montre que l'IA investit la totalité du montant disponible et sa performance est bien plus proche de la baseline qu'auparavant. Avec davantage d'entraînement, il est très probable que nous dépassions cette baseline.